

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representation of
The original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 08-339288

(43)Date of publication of application : 24.12.1996

(51)Int.Cl.

G06F 3/16

G06F 3/14

G06F 3/14

G06F 17/30

G10L 3/00

G10L 5/02

(21)Application number : 07-147291

(71)Applicant : CANON INC

(22)Date of filing : 14.06.1995

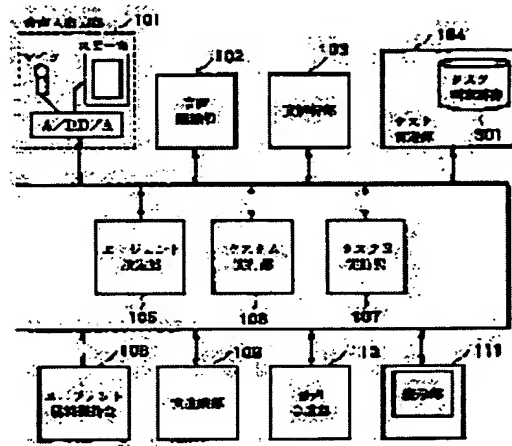
(72)Inventor : SAKAI KEIICHI
YAGISAWA TSUYOSHI
FUJITA MINORU

(54) INFORMATION PROCESSOR AND CONTROL METHOD THEREFOR

(57)Abstract:

PURPOSE: To change the form of a response sentence and the synthesized voice for each task and to make a user recognize that the task is changed in a voice interactive processing.

CONSTITUTION: In this information processor (voice interactive device) provided with the plural executable tasks, different agents are allocated corresponding to the respective plural tasks and stored in an agent attribute holding part 108. In the meantime, an input sentence recognized in a voice recognition part 102 is analyzed in an input sentence analysis part 103. A task judgment part 104 decides the task to be used based on the analyzed result and generates a command adapted to the execution task based on the input sentence. A sentence generation part 109 generates the response sentence based on a executed result obtained by executing the command by the execution task. The response sentence is formed corresponding to the attribute information of the agent corresponding to the execution task and a voice synthesis part 110 voice-synthesizes the response sentence corresponding to the attribute information.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision
of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japan Patent Office

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] The information processor which has two or more tasks which are characterized by providing the following, and which can be performed. A storing means to store different creation information corresponding to each of two or more aforementioned tasks. A determination means to determine the execution task which should analyze an input statement and should be used out of two or more aforementioned tasks. Real way stage which generates the command which is adapted for the execution task determined with the aforementioned determination means based on the aforementioned input statement, executes this command by this execution task, and obtains an execution result. A generation means to generate a response sentence based on the creation information corresponding to the aforementioned execution task stored in the aforementioned storing means, and the execution result obtained with the aforementioned execution means, and an output means to output the response sentence generated with the aforementioned generation means.

[Claim 2] The aforementioned creation information is an information processor according to claim 1 characterized by specifying the style of the response sentence generated with the aforementioned generation means.

[Claim 3] The aforementioned creation information is an information processor according to claim 1 characterized by specifying the foreground color of the response sentence generated with the aforementioned generation means.

[Claim 4] The aforementioned output means is an information processor according to claim 1 characterized by synthesizing voice based on the aforementioned response sentence, and performing a voice output.

[Claim 5] The information processor which has two or more tasks which are characterized by providing the following, and which can be performed. A storing means to store different creation information corresponding to each of two or more aforementioned tasks. A determination means to determine the execution task which should analyze an input statement and should be used out of two or more aforementioned tasks. Real way stage which generates the command which is adapted for the execution task determined with the aforementioned determination means based on the aforementioned input statement, executes this command by this execution task, and obtains an execution result. A synthetic means synthesize voice based on a generation means generate a response sentence based on the creation information corresponding to the aforementioned execution task stored in the aforementioned storing means, and the execution result obtained with the aforementioned execution means, and the creation information corresponding to the aforementioned execution task stored in the aforementioned storing means and the response sentence which were generated with the aforementioned generation means.

[Claim 6] The information processor according to claim 5 characterized by having further an input-statement generation means to generate an input statement based on the inputted sound signal.

[Claim 7] The aforementioned creation information is an information processor according to claim 5 characterized by including the information which specifies the style of the response sentence generated with the aforementioned generation means.

[Claim 8] The information processor according to claim 5 characterized by having further a display means to display the aforementioned response sentence.

[Claim 9] The aforementioned creation information is an information processor according to claim 8 characterized by specifying the foreground color by the aforementioned display means of the response sentence generated with the aforementioned generation means.

[Claim 10] The aforementioned creation information is an information processor according to claim 5 characterized by including each parameter of the speech synthesis processing in the aforementioned synthetic means. Formula.

[Claim 11] The control method of an information processor characterized by providing the following of having two or more tasks which can be performed. The storing process which stores different creation information corresponding to each of two or more aforementioned tasks. The determination process which determines the execution task which should analyze an input statement and should be used out of two or more aforementioned tasks. The execution process which generates the command which is adapted for the execution task determined at the aforementioned determination process based on the aforementioned input statement, executes this command by this execution task, and obtains an execution result. The generation process which generates a response sentence based on the creation information corresponding to the aforementioned execution task stored at the aforementioned storing process, and the execution result obtained at the aforementioned execution process, and the output process which outputs the response sentence generated at the aforementioned generation process.

[Claim 12] The control method of an information processor characterized by providing the following of having two or more tasks which can be performed. The storing process which stores different creation information corresponding to each of two or more aforementioned tasks. The determination process which determines the execution task which should analyze an input statement and should be used out of two or more aforementioned tasks. The execution process which generates the command which is adapted for the execution task determined at the aforementioned determination process based on the aforementioned input statement, executes this command by this execution task, and obtains an execution result. The synthetic process which synthesizes voice based on the generation process which generates a response sentence based on the creation information corresponding to the aforementioned execution task stored at the aforementioned storing process, and the execution result obtained at the aforementioned execution process, and the creation information corresponding to the aforementioned execution task stored at the aforementioned storing process and the response sentence which were generated at the aforementioned generation process.

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.*** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Industrial Application] this invention relates to the suitable information processor for application to voice dialog equipment, and its control method especially about the information processor which can be performed and its control method of two or more kinds of processings by changing two or more tasks.

[0002]

[Description of the Prior Art] the voice dialog equipment which can treat the task which is plurality by improvement in natural-language-processing technology in recent years and improvement in speech recognition / synthetic technology has come to be realized In this kind of voice dialog equipment, processing corresponding to the information which chose the task which should be processed based on the information inputted by voice, started, and was inputted by the started task is performed. And a synthesized speech is generated and outputted based on this processing result.

[0003]

[Problem(s) to be Solved by the Invention] However, with the above-mentioned conventional equipment, since the gestalt and synthesized speech of a response sentence which are an output are uniform, the user has not grasped whether a different task was started from the voice output. Therefore, a user cannot grasp capacity of each task of voice dialog equipment, but may demand the processing over two or more tasks. That is, the user has not grasped capacity of voice dialog equipment, but had the problem of requiring processings beyond the capacity of the voice dialog equipment concerned, such as processing which needs composite of a task.

[0004] this invention is made in view of the problem of the above-mentioned conventional technology, makes it possible to change the gestalt of the response sentence outputted for every task, and aims at offering an information processor with possible making it recognize that the task changed to the user, and its control method.

[0005] Moreover, other purposes of this invention make it possible to change the gestalt of a response sentence, and its synthesized speech for every task, and are to offer the information processor which makes it recognize that the task changed to the user in voice interactive processing, and its control method.

[0006] Moreover, it suppresses that make the capacity of each task grasp and a user performs the demand beyond the capacity of voice dialog equipment by this by making a user recognize the change rate of a task.

[0007]

[Means for Solving the Problem] The information processor of this invention for attaining the above-mentioned purpose is equipped with the following composition. Namely, a storing means to be the information processor which has two or more tasks which can be performed, and to store different creation information corresponding to each of two or more aforementioned tasks, A determination means to determine the execution task which should analyze an input statement and should be used out of two or more aforementioned tasks, The real way stage which generates the command which is adapted for the execution task determined with th

aforementioned determination means based on the aforementioned input statement, executes this command by this execution task, and obtains an execution result. It has a generation means to generate a response sentence based on the creation information corresponding to the aforementioned execution task stored in the aforementioned storing means, and the execution result obtained with the aforementioned execution means, and an output means to output the response sentence generated with the aforementioned generation means.

[0008] Moreover, the aforementioned creation information specifies preferably the style of the response sentence generated with the aforementioned generation means. For example, "** which is - measure tone" and " which are """, and "existing **" — as — a user is being able to recognize a switch of a task easily by changing a style. In this case, even if it carries out the voice output of the response sentence and carries out a display output, a user can recognize a switch of a task.

[0009] Moreover, the aforementioned creation information specifies preferably the foreground color of the response sentence generated with the aforementioned generation means. Since a switch of a task is expressed with change of a foreground color, a user can judge a switch of a task very easily.

[0010] Moreover, preferably, the aforementioned output means synthesizes voice based on the aforementioned response sentence, and performs a voice output.

[0011] Moreover, the information processor by other composition of this invention which attains the above-mentioned purpose. A storing means to be the information processor which has two or more tasks which can be performed, and to store different creation information corresponding to each of two or more aforementioned tasks. A determination means to determine the execution task which should analyze an input statement and should be used out of two or more aforementioned tasks. The real way stage which generates the command which is adapted for the execution task determined with the aforementioned determination means based on the aforementioned input statement, executes this command by this execution task, and obtains an execution result. A generation means to generate a response sentence based on the creation information corresponding to the aforementioned execution task stored in the aforementioned storing means, and the execution result obtained with the aforementioned execution means. It has a synthetic means to synthesize voice based on the creation information corresponding to the aforementioned execution task stored in the aforementioned storing means, and the response sentence generated with the aforementioned generation means.

[0012] Moreover, it has further an input-statement generation means to generate an input statement preferably based on the inputted sound signal. It is because the voice dialog equipment which can input an input statement with voice and synthesizes voice from the response sentence to this can be constituted.

[0013] Moreover, it has preferably a display means to display the aforementioned response sentence, further. A response sentence can be recognized visually and a switch of a task can be grasped more easily.

[0014] Moreover, the aforementioned creation information specifies preferably the foreground color by the aforementioned display means of the response sentence generated with the aforementioned generation means. It is because a switch can be recognized by the foreground color of a response sentence and a switch of a task can be judged more nearly intuitively.

[0015] Moreover, the aforementioned creation information contains each parameter of the speech synthesis processing in the aforementioned synthetic means preferably.

[0016]

[Function] According to the above-mentioned composition, in the information processor which has two or more tasks which can be performed, different creation information corresponding to each of two or more tasks is stored. By analyzing an input statement, the task which should be used out of two or more tasks is determined as an execution task. And the command which is adapted for the execution task based on an input statement is generated. By executing this command by the execution task, the execution result (it is equivalent to the response corresponding to an input statement) of the command concerned is obtained. And out of the creation information stored beforehand, the creation information corresponding to an execution

task is obtained, and a response sentence is generated and outputted based on this creation information and the above-mentioned execution result.

[0017] Moreover, according to other above-mentioned composition, in the information processor which has two or more tasks which can be performed, different creation information corresponding to each of two or more tasks is stored. By analyzing an input statement, the task which should be used out of two or more tasks is determined as an execution task. And the command which is adapted for the execution task based on the above-mentioned input statement is generated. It carries out by executing this command by the execution task, and an execution result is obtained. Next, the creation information corresponding to an execution task is obtained out of the creation information stored beforehand, a response sentence is generated and generated based on this creation information and an execution result, and it synthesizes voice based on creation information and a response sentence further.

[0018]

[Example] Hereafter, with reference to an attached drawing, the example of this invention is explained in detail.

[0019] Drawing 1 is the block diagram showing the block composition of the voice dialog equipment of an example. In this drawing, 101 is the voice-input/output section, while carrying out [voice / which was inputted from the microphone] A/D conversion and incorporating inside the equipment concerned, carries out D/A conversion of the generated response, and outputs it from a loudspeaker as voice. 102 is the speech recognition section, performs speech recognition to the voice inputted from the voice-input/output section 101, and generates an input statement. 103 is the sentence analysis section and performs morphological analysis, syntax analysis, and a semantic analysis to the input statement recognized in the speech recognition section 102.

[0020] 104 is the task judging section, distinguishes the task performed based on the analysis result of the input statement in the sentence analysis section 103, and generates the command which was adapted for the task to perform. The task distinction section 104 holds the task distinction dictionary 301 for distinguishing a task. In this example, in order to give explanation intelligible, two tasks shall be treated and it considers as database reference (for example, relational database which can perform reference using SQL) of extension guidance of Task A, and the full-text search (for example, thing which searches the document containing the character string made into conditions from Plain Text (poor character string)) of sightseeing guidance of Task B. Of course, the gestalt of the above-mentioned database cannot be overemphasized by that it is an example.

[0021] Drawing 3 is drawing showing the example of data composition of the task distinction dictionary 301. In the task distinction section 104, when the word registered into the task distinction dictionary 301 into the analysis result analyzed in the sentence analysis section 103 appears, a task is changed as a thing treating the task. For example, if the task judging dictionary 301 is referred to about the input statement of "wanting to know the extension of Manager Suzuki of the Administrative Division", the task A in which the word of the "Administrative Division", the "manager", and an "extension" exists will be chosen.

[0022] Consequently, the task judging section 104 will generate a command called select (extension) from(extension table) where (name = Suzuki, affiliation = Administrative Division, executive = manager) which can interpret Task A, and will pass this to Task A.

[0023] Similarly, Task B is chosen by reference of the task judging dictionary 301 to the input statement "teach the art gallery in Hakone-cho." And the task judging section 104 will generate the command ((C classification art gallery) (C address Hakone-cho) (R name ?)) which can interpret Task B, and will send this to Task B.

[0024] 105 is the agent determination section, determines an agent according to the task distinguished in the task distinction section 104, and changes an agent. 106,107 is the task A statement part and task B statement part which perform a task different, respectively. In this example, an "extension agent" presupposes that a "sightseeing agent" is chosen to Task B (sightseeing guidance), respectively to Task A (extension guidance).

[0025] 108 is an agent attribute attaching part and is an agent attribute attaching part which

holds synthesized-speech information, such as sentence generation information, such as sex and a tone, and height of voice, and phonation speed, as an attribute for every agent. Drawing 4 is drawing showing the data storage state in the agent attribute attaching part 108. Sentence generation information and synthesized-speech information are stored in each of the "extension agent" determined as the agent attribute attaching part 108 in the agent determination section, and a "sightseeing agent" as shown in drawing 4.

[0026] 109 is the sentence generation section, acquires an agent's sentence generation information determined in the agent determination section 105 from the agent attribute attaching part 108, and generates the response sentence based on the execution result of the task A statement part 106 or the task B statement part 107 with reference to this. Drawing 5 is drawing showing the example of generation of a response sentence. In drawing 5, the example of a dialog using the above-mentioned command is shown, and it is. That is, while using the extension number (here "1234") obtained as the reference conditions "Suzuki" included in the command, the "manager", an "extension", and an execution result of Task A, a response sentence is generated based on an extension agent's sentence generation information (a woman, existing **). Of course, you may make it answer only by the result like a sightseeing agent. In this case, a response sentence becomes "Being 1234." In addition, the response sentence by the sightseeing agent who outputs the execution result of Task B is the same as that of the above-mentioned extension agent.

[0027] 110 is the speech synthesis section, acquires an agent's synthesized-speech information determined in the agent determination section 105 from the agent attribute attaching part 108, and generates the synthesized speech of the response sentence generated in the sentence generation section 109 with reference to this. The synthesized speech generated in the speech synthesis section 110 is outputted as voice from the loudspeaker of the voice-input/output section 101. 111 is a display and outputs a speech recognition result, the flow of a dialog, the execution result of a task, a response sentence, etc. In the above, by the sentence generation section 109, the agent attribute of a woman and a male is referred to setting up any of language a woman talks in a response sentence, and language and a male talk, and are used, and is used for a setup of the phoneme data for generating the voice of a woman and a male in the speech synthesis section 110.

[0028] Next, operation of this equipment is explained with reference to the flow chart of drawing 2. Drawing 2 is a flow chart showing the control procedure of the voice dialog equipment of an example.

[0029] First, at Step S201, if waiting and voice input are performed in the voice input from the voice-input/output section 101, this input will be sent to the speech recognition section 102. At Step S202, the speech recognition section 102 performs speech recognition, and an input statement is obtained. Moreover, this recognition result (input statement) is displayed on a display 111, and it moves to Step S203. At Step S203, the sentence analysis section 103 performs morphological analysis, syntax analysis, and a semantic analysis to the input section which it is as a result of recognition, and moves to Step S204. At Step S204, in the task distinction section 104, the task performed from the analysis result analyzed in the sentence analysis section 103 is distinguished, and the command corresponding to the task to perform is generated.

[0030] At Step S205, an agent is determined and changed in the agent determination section 105 according to the task distinguished in the task distinction section 104. At Step S206, in the task statement part (either the task A statement part 106 or the task B statement part 107) distinguished in the task distinction section 104, the command generated at Step S204 is executed, and it moves to Step S207.

[0031] At Step S207, the sentence generation section 109 generates a response sentence from the execution result performed by the task A statement part 106 or the task B statement part 107. In generation of this response sentence, the sentence generation section 109 acquires an agent's sentence generation information determined in the agent determination section 105 from the agent attribute attaching part 108, and it generates a response sentence, referring to this sentence generation information. The generated response sentence is sent to the speech

synthesis section 110 while it is table-shown in a display 111 and is made into it.

[0032] At Step S208, the speech synthesis section 110 generates the synthesized speech of the response sentence generated in the sentence generation section 109. In generation of this synthesized speech, the speech synthesis section 110 acquires an agent's synthesized-speech information determined in the agent determination section 105 from the attribute attaching part 108, and refers to this synthesized-speech information for it. D/A conversion of the generated synthesized speech is carried out in the voice-input/output section 101, and it is outputted from a loudspeaker. Then, processing returns to Step S201.

[0033] the agent (person in charge) who stated above and who gave attributes, such as sex, a tone, height of voice, and phonation speed, like according to the voice dialog equipment of the above-mentioned example — every task — setting up — a task — changing — it becomes possible to change an agent. That is, it can be made conscious of the task having changed to the user by changing a response sentence based on sex or a tone (sentence generation information), and changing the voice of a synthesized speech based on the height and phonation speed (synthesized-speech information) of voice.

[0034] Therefore, a user becomes possible [grasping the processing range by one task], and it becomes possible to judge whether the capacity of the voice dialog equipment concerned for the demand which it is going to give from now on to need two or more tasks used to be exceeded.

[0035] In addition, although it is explained in the above-mentioned example, using the task to treat as two in order to give explanation brief, of course, it can be adapted also for the thing treating three or more tasks.

[0036] Moreover, in the above-mentioned example, as an attribute of the response sentence changed for every agent, although sex and the tone were used, it does not restrict to this. You may make it express a shift of an agent with using the color and font of the response sentence displayed on a display as an attribute of a response sentence.

[0037] Moreover, in the above-mentioned example, it does not restrict to this using the height of voice, and phonation speed as an attribute of the synthesized speech changed for every agent. You may make it express a shift of an agent with using for example, other study phoneme data, a natural descent component, pause length, etc.

[0038] Moreover, when a generation method prepares two or more mere sentence generation sections and changes the sentence generation section for every agent, you may make it generate a different response sentence, although a different response sentence by giving an attribute to the single sentence generation section is generated in the above-mentioned example.

[0039] Moreover, although a different synthesized speech by giving an attribute to the single speech synthesis section is generated in the above-mentioned example, you may make it generate a different synthesized speech by preparing two or more speech synthesis sections from which a generation method differs, and changing the synthesized-speech section for every agent.

[0040] Moreover, although both the gestalt of a response sentence and the gestalt of a synthesized speech are changed with a switch of a task, you may make it change one of gestalten in the above-mentioned example.

[0041] As explained above, according to this example, the change of a response sentence and a synthesized speech is realizable by having two or more tasks to treat and setting up recognition and the agent who does sentence analysis, judges a task from the analysis result, and changes an attribute for every task for the inputted voice.

[0042] In addition, even if it applies this invention to the system which consists of two or more devices, you may apply it to the equipment which consists of one device. Moreover, this invention cannot be overemphasized by that it can apply when attained by supplying a program to a system or equipment.

[0043]

[Effect of the Invention] As explained above, it becomes possible to change the gestalt of the response sentence outputted for every task according to this invention, and it becomes possible to make it recognize that the task changed to the user.

[0044] Moreover, according to this invention, it becomes possible to change the gestalt of a response sentence, and its synthesized speech for every task, and it becomes possible to make it recognize that the task changed to the user in voice interactive processing.

[0045] As mentioned above, it can suppress that make the capacity of each task grasp and a user performs the demand beyond the capacity of voice dialog equipment by this by making a user recognize the change rate of a task.

[0046]

[Translation done.]

*** NOTICES ***

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. *** shows the word which can not be translated.
3. In the drawings, any words are not translated.

TECHNICAL FIELD

[Industrial Application] this invention relates to the suitable information processor for application to voice dialog equipment, and its control method especially about the information processor which can be performed and its control method of two or more kinds of processings by changing two or more tasks.

[Translation done.]

*** NOTICES ***

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

PRIOR ART

[Description of the Prior Art] the voice dialog equipment which can treat the task which is plurality by improvement in natural-language-processing technology in recent years and improvement in speech recognition / synthetic technology has come to be realized In this kind of voice dialog equipment, processing corresponding to the information which chose the task which should be processed based on the information inputted by voice, started, and was inputted by the started. task is performed. And a synthesized speech is generated and outputted based on this processing result.

[Translation done.]

*** NOTICES ***

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

EFFECT OF THE INVENTION

[Effect of the Invention] As explained above, it becomes possible to change the form of the response sentence outputted for every task according to this invention, and it becomes possible to make it recognize that the task changed to the user.

[0044] Moreover, according to this invention, it becomes possible to change the form of a response sentence, and its synthesized speech for every task, and it becomes possible to make it recognize that the task changed to the user in voice interactive processing.

[0045] As mentioned above, it can suppress that make the capacity of each task grasp and a user performs the demand beyond the capacity of voice dialog equipment by this by making a user recognize the change rate of a task.

[Translation done.]

*** NOTICES ***

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. *** shows the word which can not be translated.
3. In the drawings, any words are not translated.

EFFECT OF THE INVENTION

[Effect of the Invention] As explained above, it becomes possible to change the form of the response sentence outputted for every task according to this invention, and it becomes possible to make it recognize that the task changed to the user.

[0044] Moreover, according to this invention, it becomes possible to change the form of a response sentence, and its synthesized speech for every task, and it becomes possible to make it recognize that the task changed to the user in voice interactive processing.

[0045] As mentioned above, it can suppress that make the capacity of each task grasp and a user performs the demand beyond the capacity of voice dialog equipment by this by making a user recognize the change rate of a task.

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. *** shows the word which can not be translated.
3. In the drawings, any words are not translated.

MEANS

[Means for Solving the Problem] The information processor of this invention for attaining the above-mentioned purpose is equipped with the following composition. Namely, a storing means to be the information processor which has two or more tasks which can be performed, and to store different creation information corresponding to each of two or more aforementioned tasks, A determination means to determine the execution task which should analyze an input statement and should be used out of two or more aforementioned tasks, The real way stage which generates the command which is adapted for the execution task determined with the aforementioned determination means based on the aforementioned input statement, executes this command by this execution task, and obtains an execution result, It has a generation means to generate a response sentence based on the creation information corresponding to the aforementioned execution task stored in the aforementioned storing means, and the execution result obtained with the aforementioned execution means, and an output means to output the response sentence generated with the aforementioned generation means.

[0008] Moreover, the aforementioned creation information specifies preferably the style of the response sentence generated with the aforementioned generation means. for example, ** which is "measure tone" and " which are "", and "existing **" -- as -- a user is being able to recognize a switch of a task easily by changing a style In this case, even if it carries out the voice output of the response sentence and carries out a display output, a user can recognize a switch of a task.

[0009] Moreover, the aforementioned creation information specifies preferably the foreground color of the response sentence generated with the aforementioned generation means. Since a switch of a task is expressed with change of a foreground color, a user can judge a switch of a task very easily.

[0010] Moreover, preferably, the aforementioned output means synthesizes voice based on the aforementioned response sentence, and performs a voice output.

[0011] Moreover, it is characterized by equipping with the following the information processor by other composition of this invention which attains the above-mentioned purpose. A storing means to be the information processor which has two or more tasks which can be performed, and to store different creation information corresponding to each of two or more aforementioned tasks. A determination means to determine the execution task which should analyze an input statement and should be used out of two or more aforementioned tasks. Real way stage which generates the command which is adapted for the execution task determined with the aforementioned determination means based on the aforementioned input statement, executes this command by this execution task, and obtains an execution result. A synthetic means synthesize voice based on a generation means generate a response sentence based on the creation information corresponding to the aforementioned execution task stored in the aforementioned storing means, and the execution result obtained with the aforementioned execution means, and the creation information corresponding to the aforementioned execution task stored in the aforementioned storing means and the response sentence generated with the aforementioned generation means.

[0012] Moreover, it has further an input-statement generation means to generate an input statement preferably based on the inputted sound signal. It is because the voice dialog

equipment which can input an input statement with voice and synthesizes voice from the response sentence to this can be constituted.

[0013] Moreover, it has preferably a display means to display the aforementioned response sentence, further. A response sentence can be recognized visually and a switch of a task can be grasped more easily.

[0014] Moreover, the aforementioned creation information specifies preferably the foreground color by the aforementioned display means of the response sentence generated with the aforementioned generation means. It is because a switch can be recognized by the foreground color of a response sentence and a switch of a task can be judged more nearly intuitively.

[0015] Moreover, the aforementioned creation information contains each parameter of the speech synthesis processing in the aforementioned synthetic means preferably.

[Translation done.]

*** NOTICES ***

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

OPERATION

[Function] According to the above-mentioned composition, in the information processor which has two or more tasks which can be performed, different creation information corresponding to each of two or more tasks is stored. By analyzing an input statement, the task which should be used out of two or more tasks is determined as an execution task. And the command which is adapted for the execution task based on an input statement is generated. By executing this command by the execution task, the execution result (it is equivalent to the response corresponding to an input statement) of the command concerned is obtained. And out of the creation information stored beforehand, the creation information corresponding to an execution task is obtained, and a response sentence is generated and outputted based on this creation information and the above-mentioned execution result.

[0017] Moreover, according to other above-mentioned composition, in the information processor which has two or more tasks which can be performed, different creation information corresponding to each of two or more tasks is stored. By analyzing an input statement, the task which should be used out of two or more tasks is determined as an execution task. And the command which is adapted for the execution task based on the above-mentioned input statement is generated. It carries out by executing this command by the execution task, and an execution result is obtained. Next, the creation information corresponding to an execution task is obtained out of the creation information stored beforehand, a response sentence is generated and generated based on this creation information and an execution result, and it synthesizes voice based on creation information and a response sentence further.

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.*** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

EXAMPLE

[Example] Hereafter, with reference to an attached drawing, the example of this invention is explained in detail.

[0019] Drawing 1 is the block diagram showing the block composition of the voice dialog equipment of an example. In this drawing, 101 is the voice-input/output section, while carrying out [voice / which was inputted from the microphone] A/D conversion and incorporating inside the equipment concerned, carries out D/A conversion of the generated response, and outputs it from a loudspeaker as voice. 102 is the speech recognition section, performs speech recognition to the voice inputted from the voice-input/output section 101, and generates an input statement. 103 is the sentence analysis section and performs morphological analysis, syntax analysis, and a semantic analysis to the input statement recognized in the speech recognition section 102.

[0020] 104 is the task judging section, distinguishes the task performed based on the analysis result of the input statement in the sentence analysis section 103, and generates the command which was adapted for the task to perform. The task distinction section 104 holds the task distinction dictionary 301 for distinguishing a task. In this example, in order to give explanation intelligible, two tasks shall be treated and it considers as database reference (for example, relational database which can perform reference using SQL) of extension guidance of Task A, and the full-text search (for example, thing which searches the document containing the character string made into conditions from Plain Text (poor character string)) of sightseeing guidance of Task B. Of course, the gestalt of the above-mentioned database cannot be overemphasized by that it is an example.

[0021] Drawing 3 is drawing showing the example of data composition of the task distinction dictionary 301. In the task distinction section 104, when the word registered into the task distinction dictionary 301 into the analysis result analyzed in the sentence analysis section 103 appears, a task is changed as a thing treating the task. For example, if the task judging dictionary 301 is referred to about the input statement of "wanting to know the extension of Manager Suzuki of the Administrative Division", the task A in which the word of the "Administrative Division", the "manager", and an "extension" exists will be chosen.

[0022] Consequently, the task judging section 104 will generate a command called select (extension) from(extension table) where (name = Suzuki, affiliation = Administrative Division, executive = manager) which can interpret Task A, and will pass this to Task A.

[0023] Similarly, Task B is chosen by reference of the task judging dictionary 301 to the input statement "teach the art gallery in Hakone-cho." And the task judging section 104 will generate the command ((C classification art gallery) (C address Hakone-cho) (R name ?)) which can interpret Task B, and will send this to Task B.

[0024] 105 is the agent determination section, determines an agent according to the task distinguished in the task distinction section 104, and changes an agent. 106,107 is the task A statement part and task B statement part which perform a task different, respectively. In this example, an "extension agent" presupposes that a "sightseeing agent" is chosen to Task B (sightseeing guidance), respectively to Task A (extension guidance).

[0025] 108 is an agent attribute attaching part and is an agent attribute attaching part which

holds synthesized-speech information, such as sentence generation information, such as sex and a tone, and height of voice, and phonation speed, as an attribute for every agent. Drawing 4 is drawing showing the data storage state in the agent attribute attaching part 108. Sentence generation information and synthesized-speech information are stored in each of the "extension agent" determined as the agent attribute attaching part 108 in the agent determination section, and a "sightseeing agent" as shown in drawing 4.

[0026] 109 is the sentence generation section, acquires an agent's sentence generation information determined in the agent determination section 105 from the agent attribute attaching part 108, and generates the response sentence based on the execution result of the task A statement part 106 or the task B statement part 107 with reference to this. Drawing 5 is drawing showing the example of generation of a response sentence. In drawing 5, the example of a dialog using the above-mentioned command is shown, and it is. That is, while using the extension number (here "1234") obtained as the reference conditions "Suzuki" included in the command, the "manager", an "extension", and an execution result of Task A, a response sentence is generated based on an extension agent's sentence generation information (a woman, existing **). Of course, you may make it answer only by the result like a sightseeing agent. In this case, a response sentence becomes "Being 1234." In addition, the response sentence by the sightseeing agent who outputs the execution result of Task B is the same as that of the above-mentioned extension agent.

[0027] 110 is the speech synthesis section, acquires an agent's synthesized-speech information determined in the agent determination section 105 from the agent attribute attaching part 108, and generates the synthesized speech of the response sentence generated in the sentence generation section 109 with reference to this. The synthesized speech generated in the speech synthesis section 110 is outputted as voice from the loudspeaker of the voice-input/output section 101. 111 is a display and outputs a speech recognition result, the flow of a dialog, the execution result of a task, a response sentence, etc. In the above, by the sentence generation section 109, the agent attribute of a woman and a male is referred to setting up any of language a woman talks in a response sentence, and language and a male talk, and are used, and is used for a setup of the phoneme data for generating the voice of a woman and a male in the speech synthesis section 110.

[0028] Next, operation of this equipment is explained with reference to the flow chart of drawing 2. Drawing 2 is a flow chart showing the control procedure of the voice dialog equipment of an example.

[0029] First, at Step S201, if waiting and voice input are performed in the voice input from the voice-input/output section 101, this input will be sent to the speech recognition section 102. At Step S202, the speech recognition section 102 performs speech recognition, and an input statement is obtained. Moreover, this recognition result (input statement) is displayed on a display 111, and it moves to Step S203. At Step S203, the sentence analysis section 103 performs morphological analysis, syntax analysis, and a semantic analysis to the input section which it is as a result of recognition, and moves to Step S204. At Step S204, in the task distinction section 104, the task performed from the analysis result analyzed in the sentence analysis section 103 is distinguished, and the command corresponding to the task to perform is generated.

[0030] At Step S205, an agent is determined and changed in the agent determination section 105 according to the task distinguished in the task distinction section 104. At Step S206, in the task statement part (either the task A statement part 106 or the task B statement part 107) distinguished in the task distinction section 104, the command generated at Step S204 is executed, and it moves to Step S207.

[0031] At Step S207, the sentence generation section 109 generates a response sentence from the execution result performed by the task A statement part 106 or the task B statement part 107. In generation of this response sentence, the sentence generation section 109 acquires an agent's sentence generation information determined in the agent determination section 105 from the agent attribute attaching part 108, and it generates a response sentence, referring to this sentence generation information. The generated response sentence is sent to the speech

synthesis section 110 while it is table-shown in a display 111 and is mad into it.

[0032] At Step S208, the speech synthesis section 110 generates the synthesized speech of the response sentence generated in the sentence generation section 109. In generation of this synthesized speech, the speech synthesis section 110 acquires an agent's synthesized-speech information determined in the agent determination section 105 from the attribute attaching part 108, and refer to this synthesized-speech information for it. D/A conversion of the generated synthesized speech is carried out in the voice-input/output section 101, and it is outputted from a loudspeaker. Then, processing returns to Step S201.

[0033] the agent (person in charge) who stated above and who gave attributes, such as sex, a tone, height of voice, and phonation speed, like according to the voice dialog equipment of the above-mentioned example — every task — setting up — a task — changing — it becomes possible to change an agent. That is, it can be made conscious of the task having changed to the user by changing a response sentence based on sex or a tone (sentence generation information), and changing the voice of a synthesized speech based on the height and phonation speed (synthesized-speech information) of voice.

[0034] Therefore, a user becomes possible [grasping the processing range by one task], and it becomes possible to judge whether the capacity of the voice dialog equipment concerned for the demand which it is going to give from now on to need two or more tasks used to be exceeded.

[0035] In addition, although it is explained in the above-mentioned example, using the task to treat as two in order to give explanation brief, of course, it can be adapted also for the thing treating three or more tasks.

[0036] Moreover, in the above-mentioned example, as an attribute of the response sentence changed for every agent, although sex and the tone were used, it does not restrict to this. You may make it express a shift of an agent with using the color and font of the response sentence displayed on a display as an attribute of a response sentence.

[0037] Moreover, in the above-mentioned example, it does not restrict to this using the height of voice, and phonation speed as an attribute of the synthesized speech changed for every agent. You may make it express a shift of an agent with using for example, other study phoneme data, a natural descent component, pause length, etc.

[0038] Moreover, when a generation method prepares two or more mere sentence generation sections and changes the sentence generation section for every agent, you may make it generate a different response sentence, although a different response sentence by giving an attribute to the single sentence generation section is generated in the above-mentioned example.

[0039] Moreover, although a different synthesized speech by giving an attribute to the single speech synthesis section is generated in the above-mentioned example, you may make it generate a different synthesized speech by preparing two or more speech synthesis sections from which a generation method differs, and changing the synthesized-speech section for every agent.

[0040] Moreover, although both the gestalt of a response sentence and the gestalt of a synthesized speech are changed with a switch of a task, you may make it change one of gestalten in the above-mentioned example.

[0041] As explained above, according to this example, the change of a response sentence and a synthesized speech is realizable by having two or more tasks to treat and setting up recognition and the agent who does sentence analysis, judges a task from the analysis result, and changes an attribute for every task for the inputted voice.

[0042] In addition, even if it applies this invention to the system which consists of two or more devices, you may apply it to the equipment which consists of one device. Moreover, this invention cannot be overemphasized by that it can apply when attained by supplying a program to a system or equipment.

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.*** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] It is drawing showing the block composition of the voice dialog equipment of an example.

[Drawing 2] It is a flow chart showing the control procedure of the voice dialog equipment of an example.

[Drawing 3] It is drawing showing the example of data composition of the task distinction dictionary 301.

[Drawing 4] It is drawing showing the data storage state in the agent attribute attaching part 108.

[Drawing 5] It is drawing showing the example of generation of a response sentence.

[Description of Notations]

101 Voice-Input/output Section

102 Speech Recognition Section

103 Sentence Analysis Section

104 Task Distinction Section

105 Agent Determination Section

106 Task A Statement Part

107 Task B Statement Part

108 Agent Attribute Attaching Part

109 Sentence Generation Section

110 Speech Synthesis Section

111 Display

[Translation done.]

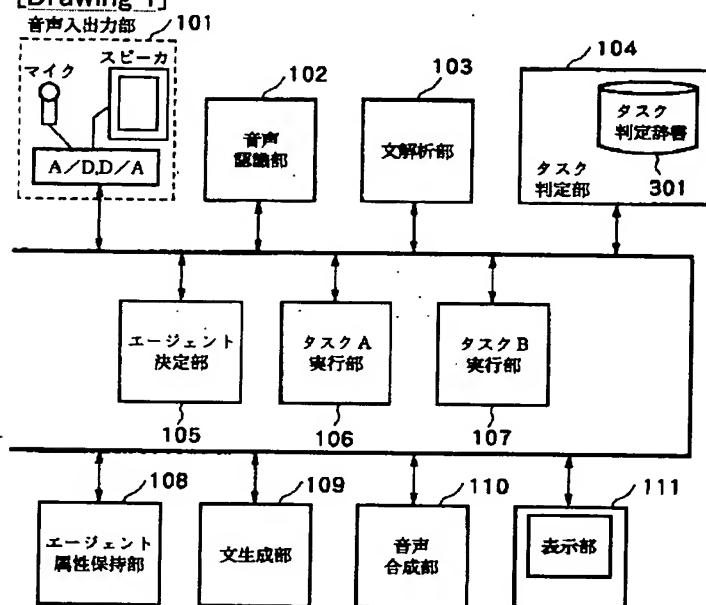
* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

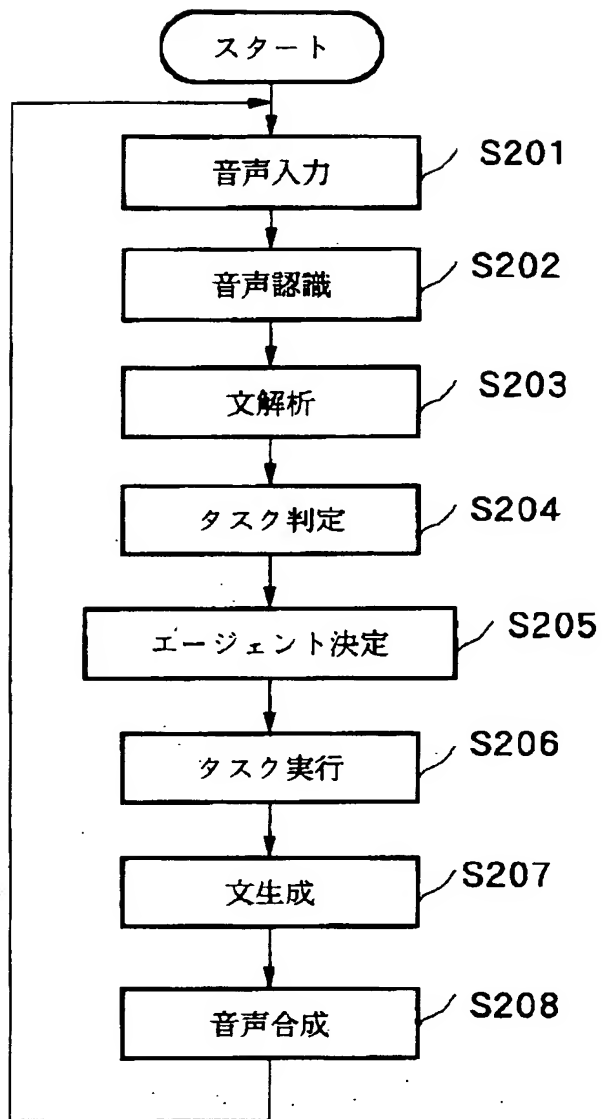
1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DRAWINGS

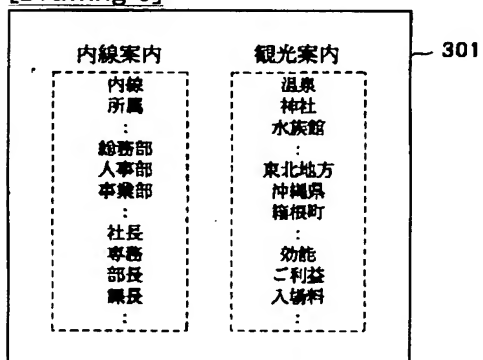
[Drawing 1]



[Drawing 2]



[Drawing 3]



[Drawing 4]

108

内線 エージェント		観光 エージェント	
性別	女性	性別	男性
語調	ございます調	語調	ですます調
声の高さ	880	声の高さ	660
発声速度	4	発声速度	3
：	：	：	：

[Drawing 5]

ユーザ	総務部の鈴木部長の内線を知りたい。
内線エージェント	鈴木部長の内線は1234でございます。
ユーザ	箱根町にある美術館を教えてください。
観光エージェント	彫刻の森美術館,成川美術館,箱根仙石原美術館,箱根美術館です。

[Translation done.]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平8-339288

(43) 公開日 平成8年(1996)12月24日

(51) Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 3/16	3 3 0	9172-5E	G 0 6 F 3/16	3 3 0 H
3/14	3 3 0		3/14	3 3 0 A
	3 4 0			3 4 0 A
17/30			G 1 0 L 3/00	5 7 1 H
G 1 0 L 3/00	5 7 1		5/02	J
審査請求 未請求 請求項の数12 O L (全 7 頁) 最終頁に続く				

(21) 出願番号 特願平7-147291

(22) 出願日 平成7年(1995)6月14日

(71) 出願人 000001007

キヤノン株式会社

東京都大田区下丸子3丁目30番2号

(72) 発明者 酒井 桂一

東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

(72) 発明者 八木沢 津義

東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

(72) 発明者 藤田 稔

東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

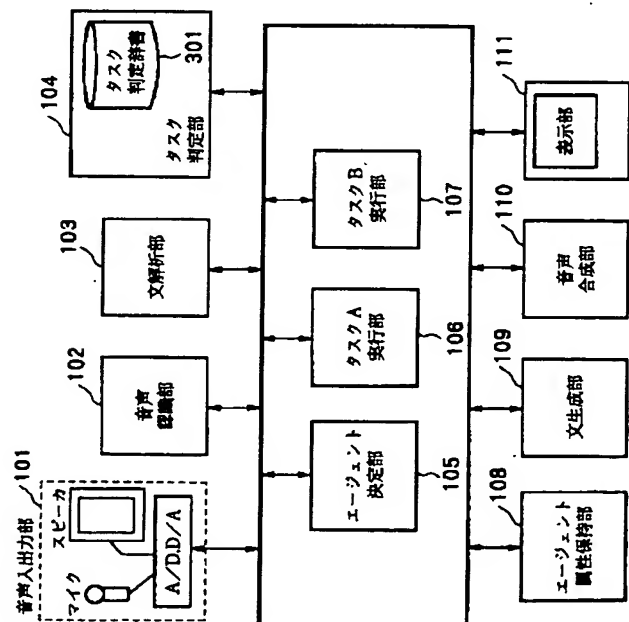
(74) 代理人 弁理士 大塚 康徳 (外1名)

(54) 【発明の名称】 情報処理装置及びその制御方法

(57) 【要約】

【目的】 応答文の形態およびその合成音声をタスク毎に変化させることを可能とし、音声対話処理において利用者にタスクが変わったことを認識させる。

【構成】 実行可能な複数のタスクを有する情報処理装置（音声対話装置）において、複数のタスクのそれぞれに対応して異なるエージェントが割り当てられ、エージェント属性保持部108に格納される。一方、音声認識部102で認識された入力文は入力文解析部103で解析される。タスク判定部104はこの解析結果に基づいて使用すべきタスクを決定し、上記入力文に基づく、実行タスクに適応するコマンドを生成する。文生成部109は、このコマンドを実行タスクで実行して得られた実行結果に基づく応答文を生成する。この応答文は、当該実行タスクに対応するエージェントの属性情報に従って形成されるとともに、音声合成部110はこの応答文を上記属性情報に従って音声合成する。



【特許請求の範囲】

【請求項1】 実行可能な複数のタスクを有する情報処理装置であって、

前記複数のタスクのそれぞれに対応して異なる生成情報を格納する格納手段と、

入力文を解析して前記複数のタスクの中から使用すべき実行タスクを決定する決定手段と、

前記入力文に基づいて前記決定手段で決定された実行タスクに適応するコマンドを生成し、該コマンドを該実行タスクで実行して実行結果を得る実行手段と、

前記格納手段に格納されている前記実行タスクに対応する生成情報と前記実行手段で得られた実行結果とに基づいて応答文を生成する生成手段と、

前記生成手段で生成された応答文を出力する出力手段とを備えることを特徴とする情報処理装置。

【請求項2】 前記生成情報は、前記生成手段で生成される応答文の文体を規定することを特徴とする請求項1に記載の情報処理装置。

【請求項3】 前記生成情報は、前記生成手段で生成される応答文の表示色を規定することを特徴とする請求項1に記載の情報処理装置。

【請求項4】 前記出力手段は、前記応答文に基づいて音声合成を行い、音声出力を行うことを特徴とする請求項1に記載の情報処理装置。

【請求項5】 実行可能な複数のタスクを有する情報処理装置であって、

前記複数のタスクのそれぞれに対応して異なる生成情報を格納する格納手段と、

入力文を解析して前記複数のタスクの中から使用すべき実行タスクを決定する決定手段と、

前記入力文に基づいて前記決定手段で決定された実行タスクに適応するコマンドを生成し、該コマンドを該実行タスクで実行して実行結果を得る実行手段と、

前記格納手段に格納されている前記実行タスクに対応する生成情報と前記実行手段で得られた実行結果とに基づいて応答文を生成する生成手段と、

前記格納手段に格納されている前記実行タスクに対応する生成情報と前記生成手段で生成された応答文とに基づいて音声合成する合成手段とを備えることを特徴とする情報処理装置。

【請求項6】 入力された音声信号に基づいて入力文を生成する入力文生成手段を更に備えることを特徴とする請求項5に記載の情報処理装置。

【請求項7】 前記生成情報は、前記生成手段で生成される応答文の文体を規定する情報を含むことを特徴とする請求項5に記載の情報処理装置。

【請求項8】 前記応答文を表示する表示手段を更に備えることを特徴とする請求項5に記載の情報処理装置。

【請求項9】 前記生成情報は、前記生成手段で生成される応答文の前記表示手段による表示色を規定すること

を特徴とする請求項8に記載の情報処理装置。

【請求項10】 前記生成情報は、前記合成手段における音声合成処理の各パラメータを含むことを特徴とする請求項5に記載の情報処理装置。式。

【請求項11】 実行可能な複数のタスクを有する情報処理装置の制御方法であって、

前記複数のタスクのそれぞれに対応して異なる生成情報を格納する格納工程と、

入力文を解析して前記複数のタスクの中から使用すべき実行タスクを決定する決定工程と、

前記入力文に基づいて前記決定工程で決定された実行タスクに適応するコマンドを生成し、該コマンドを該実行タスクで実行して実行結果を得る実行工程と、

前記格納工程で格納された前記実行タスクに対応する生成情報と前記実行工程で得られた実行結果とに基づいて応答文を生成する生成工程と、

前記生成工程で生成された応答文を出力する出力工程とを備えることを特徴とする制御方法。

【請求項12】 実行可能な複数のタスクを有する情報処理装置の制御方法であって、

前記複数のタスクのそれぞれに対応して異なる生成情報を格納する格納工程と、

入力文を解析して前記複数のタスクの中から使用すべき実行タスクを決定する決定工程と、

前記入力文に基づいて前記決定工程で決定された実行タスクに適応するコマンドを生成し、該コマンドを該実行タスクで実行して実行結果を得る実行工程と、

前記格納工程にて格納された前記実行タスクに対応する生成情報と前記実行工程で得られた実行結果とに基づいて応答文を生成する生成工程と、

前記格納工程にて格納された前記実行タスクに対応する生成情報と前記生成工程で生成された応答文とに基づいて音声合成する合成工程とを備えることを特徴とする制御方法。

【発明の詳細な説明】**【0001】**

【産業上の利用分野】 本発明は複数のタスクを切り替えることで複数種類の処理の実行が可能な情報処理装置及びその制御方法に関し、特に、音声対話装置への適用に好適な情報処理装置及びその制御方法に関するものである。

【0002】

【従来の技術】 近年の自然言語処理技術の向上、音声認識／合成技術の向上により、複数のタスクを扱える音声対話装置が実現されるようになってきた。この種の音声対話装置においては、音声によって入力された情報に基づいて処理すべきタスクを選択、起動し、起動されたタスクにより入力された情報に対応する処理が実行される。そして、この処理結果に基づいて、合成音声が生産され、出力される。

【0003】

【発明が解決しようとする課題】しかしながら、上記従来の装置では、出力である応答文の形態および合成音声が一面的であるため、利用者は、異なるタスクが起動されたか否かをその音声出力から把握することはできなかった。従って、利用者が音声対話装置の各タスクの能力を把握できず、複数のタスクにまたがった処理を要求してしまう可能性がある。即ち、利用者は、音声対話装置の能力を把握できず、タスクの複合を必要とする処理など、当該音声対話装置の能力以上の処理を要求をしてしまうという問題があった。

【0004】本発明は、上記従来技術の問題に鑑みてなされたものであり、タスク毎に出力される応答文の形態を変化させることを可能とし、利用者にタスクが変わったことを認識させることが可能な情報処理装置及びその制御方法を提供することを目的とする。

【0005】また、本発明の他の目的は、応答文の形態およびその合成音声をタスク毎に変化させることを可能とし、音声対話処理において利用者にタスクが変わったことを認識させる情報処理装置及びその制御方法を提供することにある。

【0006】また、利用者にタスクの切り替わりを認識させることで、各タスクの能力を把握させ、これにより、利用者が音声対話装置の能力以上の要求を行うことを抑制する。

【0007】

【課題を解決するための手段】上記の目的を達成するための本発明の情報処理装置は以下の構成を備えている。即ち、実行可能な複数のタスクを有する情報処理装置であって、前記複数のタスクのそれぞれに対応して異なる生成情報を格納する格納手段と、入力文を解析して前記複数のタスクの中から使用すべき実行タスクを決定する決定手段と、前記入力文に基づいて前記決定手段で決定された実行タスクに適応するコマンドを生成し、該コマンドを該実行タスクで実行して実行結果を得る実行手段と、前記格納手段に格納されている前記実行タスクに対応する生成情報と前記実行手段で得られた実行結果とに基づいて応答文を生成する生成手段と、前記生成手段で生成された応答文を出力する出力手段とを備える。

【0008】また、好ましくは、前記生成情報は、前記生成手段で生成される応答文の文体を規定する。例えば、「です・ます調」、「である調」、「ございます調」というように文体を変更することで、利用者は容易にタスクの切り換えを認識できるようになる。この場合、応答文を音声出力しても表示出力しても利用者はタスクの切り換えを認識できる。

【0009】また、好ましくは、前記生成情報は、前記生成手段で生成される応答文の表示色を規定する。表示色の変化でタスクの切り換えを表すので、利用者は極めて容易にタスクの切り換えを判断できる。

【0010】また、好ましくは、前記出力手段は、前記応答文に基づいて音声合成を行い、音声出力を行う。

【0011】また、上記の目的を達成する本発明の他の構成による情報処理装置は、実行可能な複数のタスクを有する情報処理装置であって、前記複数のタスクのそれぞれに対応して異なる生成情報を格納する格納手段と、入力文を解析して前記複数のタスクの中から使用すべき実行タスクを決定する決定手段と、前記入力文に基づいて前記決定手段で決定された実行タスクに適応するコマンドを生成し、該コマンドを該実行タスクで実行して実行結果を得る実行手段と、前記格納手段に格納されている前記実行タスクに対応する生成情報と前記実行手段で得られた実行結果とに基づいて応答文を生成する生成手段と、前記格納手段に格納されている前記実行タスクに対応する生成情報と前記生成手段で生成された応答文とに基づいて音声合成する合成手段とを備える。

【0012】また、好ましくは、入力された音声信号に基づいて入力文を生成する入力文生成手段を更に備える。音声によって入力文を入力でき、これに対する応答文を音声合成する音声対話装置を構成できるからである。

【0013】また、好ましくは、前記応答文を表示する表示手段を更に備える。応答文を視覚的に認識でき、タスクの切り換えをより容易に把握できる。

【0014】また、好ましくは、前記生成情報は、前記生成手段で生成される応答文の前記表示手段による表示色を規定する。応答文の表示色で切り換えを認識でき、より直観的にタスクの切り換えを判断できるからである。

【0015】また、好ましくは、前記生成情報は、前記合成手段における音声合成処理の各パラメータを含む。

【0016】

【作用】上記の構成によれば、実行可能な複数のタスクを有する情報処理装置において、複数のタスクのそれぞれに対応して異なる生成情報が格納される。入力文を解析することにより、複数のタスクの中から使用すべきタスクが実行タスクとして決定される。そして、入力文に基づく、実行タスクに適応するコマンドが生成される。このコマンドを実行タスクで実行することにより、当該コマンドに対する実行結果（入力文に対応する応答に相当する）を得る。そして、予め格納されている生成情報の中から、実行タスクに対応する生成情報を得て、この生成情報と上記実行結果とに基づいて応答文を生成し、出力する。

【0017】また、上記の他の構成によれば、実行可能な複数のタスクを有する情報処理装置において、複数のタスクのそれぞれに対応して異なる生成情報が格納される。入力文を解析することにより、複数のタスクの中から使用すべきタスクが実行タスクとして決定される。そして、上記入力文に基づく、実行タスクに適応するコマ

ンドが生成される。このコマンドを実行タスクで実行することによりして実行結果を得る。次に、予め格納されている生成情報の中から実行タスクに対応する生成情報を得て、この生成情報と実行結果とに基づいて応答文を生成する生成し、更に生成情報と応答文に基づいて音声合成を行う。

【0018】

【実施例】以下、添付の図面を参照して本発明の実施例を詳細に説明する。

【0019】図1は実施例の音声対話装置のブロック構成を示すブロック図である。同図において、101は音声入出力部であり、マイクより入力された音声をA/D変換して当該装置内部に取り込む一方、生成された応答をD/A変換して、音声としてスピーカより出力する。102は音声認識部であり、音声入出力部101より入力された音声に対して音声認識を行い、入力文を生成する。103は文解析部であり、音声認識部102において認識された入力文に対し、形態素解析、構文解析、意味解析を行う。

【0020】104はタスク判定部であり、文解析部103における入力文の解析結果に基づいて実行するタスクを判別し、実行するタスクに適応したコマンドを生成する。タスク判別部104は、タスクを判別するためのタスク判別辞書301を保持している。本実施例では、説明をわかりやすくするため2つのタスクを扱うものとし、タスクAを内線案内のデータベース検索（例えば、SQLを用いて検索の行えるリレーショナルデータベース）、タスクBを観光案内の全文検索（例えばプレーンテキスト（べた文字列）から条件とした文字列を含む文書を検索するもの）とする。もちろん上記のデータベースの形態は一例であることは、いうまでもない。

【0021】図3はタスク判別辞書301のデータ構成例を示す図である。タスク判別部104では、文解析部103で解析された解析結果中にタスク判別辞書301に登録されている語が現れた場合に、そのタスクを扱うものとして、タスクを切り替える。例えば、「総務部の鈴木部長の内線を知りたい。」という入力文についてタスク判定辞書301を参照すると、「総務部」、「部長」、「内線」という語が存在するタスクAが選択される。

【0022】この結果、タスク判定部104は、タスクAが解釈可能な、select（内線）from（内線テーブル）where（名称＝鈴木、所属＝総務部、役職＝部長）というコマンドを生成し、これをタスクAに渡すことになる。

【0023】同様に、「箱根町にある美術館を教えてください。」という入力文に対しては、タスク判定辞書301の参照によりタスクBが選択される。そして、タスク判定部104はタスクBが解釈可能な、（（C分類 美術館）（C所在地 箱根町）（R 名称？））というコマ

ンドを生成し、これをタスクBに送ることになる。

【0024】105はエージェント決定部であり、タスク判別部104で判別されたタスクに応じてエージェントを決定し、エージェントの切り替えを行う。106、107は、それぞれ異なるタスクを実行するタスクA実行部、タスクB実行部である。本例では、タスクA（内線案内）に対しては「内線エージェント」が、タスクB（観光案内）に対しては「観光エージェント」が夫々選択されるとする。

【0025】108はエージェント属性保持部であり、性別、語調などの文生成情報、及び声の高さ、発声速度などの合成音声情報をエージェントごとに属性として保持するエージェント属性保持部である。図4はエージェント属性保持部108におけるデータ格納状態を表す図である。図4に示されているように、エージェント属性保持部108には、エージェント決定部で決定される「内線エージェント」と「観光エージェント」の夫々に、文生成情報及び合成音声情報が格納されている。

【0026】109は文生成部であり、エージェント決定部105で決定されたエージェントの文生成情報をエージェント属性保持部108より獲得し、これを参照してタスクA実行部106或はタスクB実行部107の実行結果に基づく応答文を生成する。図5は応答文の生成例を表す図である。図5では、上述のコマンドを用いた対話例が示されている。即ち、コマンドに含まれている検索条件「鈴木」、「部長」、「内線」と、タスクAの実行結果として得られた内線番号（ここでは「1234」）を用いるとともに、内線エージェントの文生成情報（女性、ございます調）に基づいて応答文が生成される。もちろん、観光エージェントのように結果のみで応答するようにしても良い。この場合、応答文は、「1234でございます」となる。なお、タスクBの実行結果を出力する観光エージェントによる応答文も、上記内線エージェントと同様である。

【0027】110は音声合成部であり、エージェント決定部105で決定されたエージェントの合成音声情報をエージェント属性保持部108より獲得し、これを参照して、文生成部109で生成された応答文の合成音声を生成する。音声合成部110で生成された合成音声は音声入出力部101のスピーカより音声として出力される。111は表示部であり、音声認識結果、対話の流れ、タスクの実行結果、応答文などを出力する。上記において、女性、男性のエージェント属性は、文生成部109では、応答文において女性の話し言葉、男性の話し言葉のいずれを用いるかを設定するのに参照され、音声合成部110では女性、男性の声を生成するための音素データの設定に用いられる。

【0028】次に図2のフローチャートを参照して、本装置の動作を説明する。図2は実施例の音声対話装置の制御手順を表すフローチャートである。

【0029】まず、ステップS201では、音声入出力部101よりの音声入力待ち、音声入力が行われると、この入力情報を音声認識部102に送る。ステップS202では、音声認識部102により音声認識を行い入力文を得る。また、この認識結果（入力文）を表示部111に表示してステップS203に移る。ステップS203では、文解析部103が、認識結果である入力部に対して形態素解析、構文解析、意味解析を行い、ステップS204に移る。ステップS204では、タスク判別部104において、文解析部103で解析された解析結果から実行するタスクを判別し、実行するタスクに対応したコマンドを生成する。

【0030】ステップS205では、エージェント決定部105において、タスク判別部104で判別されたタスクに応じてエージェントを決定し、切り替える。ステップS206では、タスク判別部104で判別されたタスク実行部（タスクA実行部106あるいはタスクB実行部107のいずれか）において、ステップS204で生成されたコマンドを実行し、ステップS207に移る。

【0031】ステップS207では、文生成部109が、タスクA実行部106あるいはタスクB実行部107で実行された実行結果から応答文を生成する。この応答文の生成にあたって、文生成部109はエージェント決定部105で決定されたエージェントの文生成情報をエージェント属性保持部108より獲得し、この文生成情報を参照しながら応答文を生成する。生成された応答文は表示部111に表示されるとともに、音声合成部110に送られる。

【0032】ステップS208では、音声合成部110が、文生成部109で生成された応答文の合成音声生成する。この合成音声の生成にあたって、音声合成部110はエージェント決定部105で決定されたエージェントの合成音声情報を属性保持部108より獲得し、この合成音声情報を参照する。生成された合成音声は音声入出力部101でD/A変換されてスピーカより出力される。その後、処理はステップS201へ戻る。

【0033】以上述べたように、上記実施例の音声対話装置によれば、性別、語調、声の高さ、発声速度などの属性を与えたエージェント（担当者）をタスクごとに設定し、タスクの切り替わりによってエージェントを切り替えることが可能となる。すなわち、性別や語調（文生成情報）に基づいて応答文を切り替え、声の高さや発声速度（合成音声情報）に基づいて合成音声の音声を切り替えることで、ユーザにタスクが変わったことを意識させることができる。

【0034】従ってユーザは1つのタスクによる処理範囲を把握することが可能となり、これから与えようとする要求が複数のタスクを必要とするような当該音声対話装置の能力を超えたものであるか否かを判断することが

可能となる。

【0035】なお、上記実施例では、説明を簡潔にするために、扱うタスクを2つとして説明しているが、もちろん、3つ以上のタスクを扱うものにも適応できるものである。

【0036】また、上記実施例では、エージェントごとに切り替える応答文の属性として、性別、語調を用いたがこれに限らない。応答文の属性として、例えば表示部に表示する応答文の色や字体を用いることで、エージェントの交代を表すようにしても構わない。

【0037】また、上記実施例では、エージェントごとに切り替える合成音声の属性として、声の高さ、発声速度を用いたこれに限らない。他の例えば、学習音素データや自然降下成分、ポーズ長などを用いることで、エージェントの交代を表すようにしても構わない。

【0038】また、上記実施例では、単一の文生成部に対して属性を与えることによって異なる応答文を生成しているが、生成方式が単なる複数の文生成部を用意し、エージェントごとに文生成部を切り替えることにより、異なる応答文を生成するようにしても構わない。

【0039】また、上記実施例では、単一の音声合成部に対して属性を与えることによって異なる合成音声生成しているが、生成方式が異なる複数の音声合成部を用意し、エージェントごとに音声合成部を切り替えることにより、異なる合成音声生成するようにしても構わない。

【0040】また、上記実施例では、タスクの切り換えに伴って応答文の形態と合成音声の形態の両方を変化させているが、何れか一方の形態を変化させるようにしても良い。

【0041】以上説明したように、本実施例によれば、扱うタスクを複数持ち、入力された音声認識、文解析し、その解析結果からタスクを判定し、タスクごとに属性を切り替えるエージェントを設定することで、応答文および合成音声の切り替えを実現できる。

【0042】尚、本発明は、複数の機器から構成されるシステムに適用しても、1つの機器から成る装置に適用しても良い。また、本発明はシステム或は装置にプログラムを供給することによって達成される場合にも適用できることはいうまでもない。

【0043】

【発明の効果】以上説明したように本発明によれば、タスク毎に出力される応答文の形態を変化させることが可能となり、利用者にタスクが変わったことを認識させることが可能となる。

【0044】また、本発明によれば、応答文の形態およびその合成音声をタスク毎に変化させることが可能となり、音声対話処理において利用者にタスクが変わったことを認識させることが可能となる。

【0045】以上のように、利用者にタスクの切り替わ

りを認識させることで、各タスクの能力を把握させ、これにより、利用者が音声対話装置の能力以上の要求を行うことを抑制できる。

【0046】

【図面の簡単な説明】

【図1】実施例の音声対話装置のブロック構成を示す図である。

【図2】実施例の音声対話装置の制御手順を表すフローチャートである。

【図3】タスク判別辞書301のデータ構成例を示す図である。

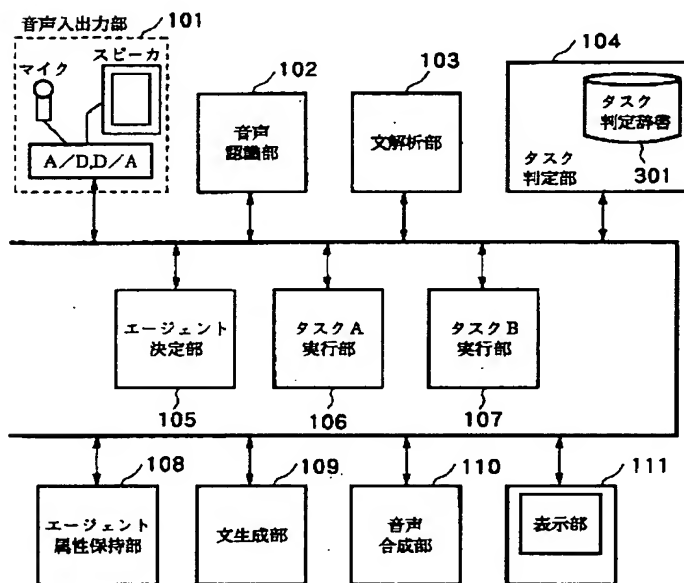
【図4】エージェント属性保持部108におけるデータ格納状態を表す図である。

【図5】応答文の生成例を表す図である。

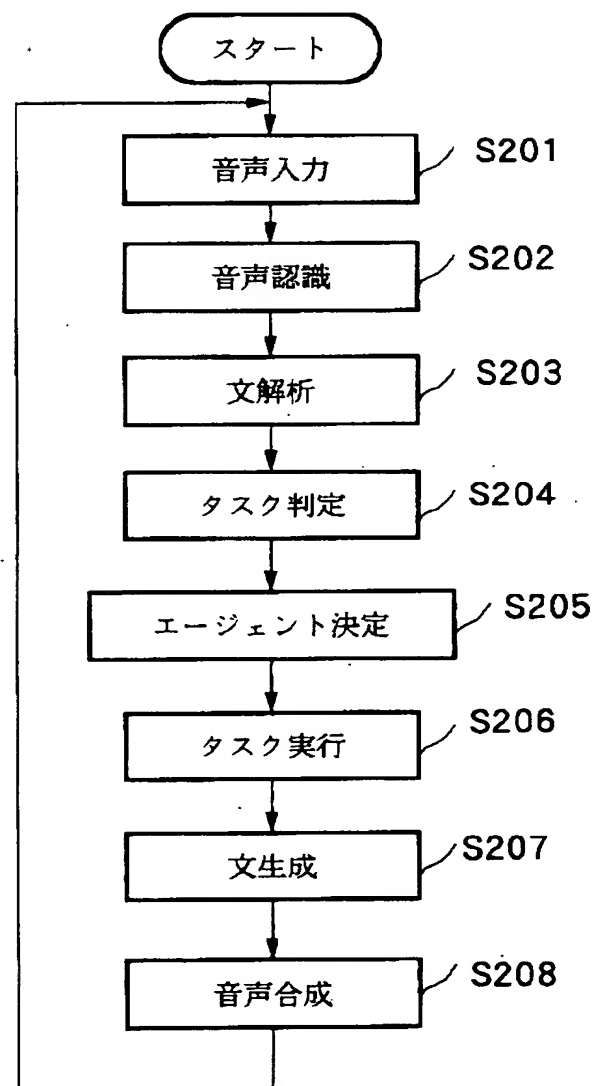
【符号の説明】

- 101 音声入出力部
- 102 音声認識部
- 103 文解析部
- 104 タスク判別部
- 105 エージェント決定部
- 106 タスクA実行部
- 107 タスクB実行部
- 108 エージェント属性保持部
- 109 文生成部
- 110 音声合成部
- 111 表示部

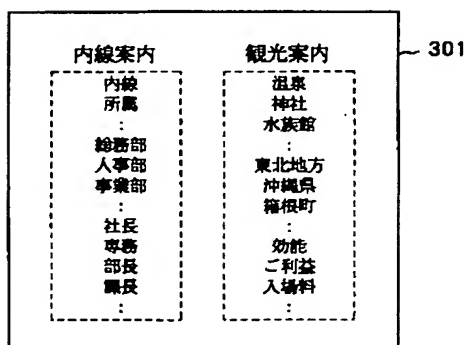
【図1】



【図2】



【図3】



【図4】

内線 エージェント		観光 エージェント	
性別	女性	性別	男性
語調	ございます調	語調	ですます調
声の高さ	880	声の高さ	660
発声速度	4	発声速度	3

【図5】

ユーザ	総務部の鈴木部長の内線を知りたい。
内線エージェント	鈴木部長の内線は1234でございます。
ユーザ	箱根町にある美術館を教えてください。
観光エージェント	彫刻の森美術館、成川美術館、箱根仙石原美術館、箱根美術館です。

・ フロントページの続き

(51) Int. Cl. 6

G 1 0 L 5/02

識別記号

庁内整理番号

9289-5 L

F I

G 0 6 F 15/403

技術表示箇所

3 1 0 Z